Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

13

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 82-1090	2. GOVT ACCESSION NO.: AD-A123 147
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Determining Behavioral Consultation Effectiveness in The United States Air Force	5. TYPE OF REPORT & PERIOD COVERED: Final
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Edward J. Conlon, University of Iowa	8. CONTRACT OR GRANT NUMBER(s): AFOSR-81-0138
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Georgia Institute of Technology, Atlanta, Ga. 30332	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2313/D9
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research (NL), Bolling AFB DC 20332	12. REPORT DATE: 10-12-1982
	13. NUMBER OF PAGES: 38
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution unlimited

Approved for public release; distribution unlimited.

DTIC ELECTE JAN 7 1983 S D A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Prepared in cooperation with: Leadership and Management Development Center, Maxwell AFB, AL 36112

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Organizational Change
Organizational Development
Survey Feedback
Evaluation Research

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

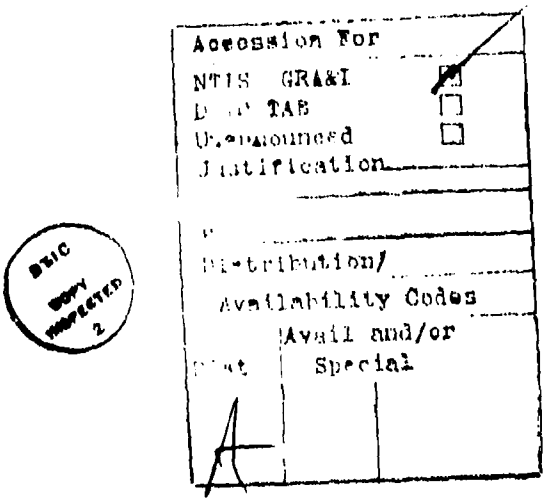
This report describes an analysis of evaluation data collected as part of The Air Force Leadership and Management Development Centers program of Survey Feedback. Several analyses were performed to determine the impact that consultation was having on attitudes and perceptions within the Air Force. The results indicate that although consultation seems to have some effect, the processes mediating those effects are unclear.

DD FORM 1 JAN 73 1473

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Determining Behavioral Consultation Effectiveness in the United States Air Force

Edward J. Conlon Ph.D.
College of Management
Georgia Institute of Technology
Atlanta, GA 30332
(404) 894-2612



1mlco/10/6/82

This report describes a study which is an extention of research conducted under the Air Force Office of Scientific Research's (AFOSR) Summer Faculty Research Program (SFRP). It is submitted in completion of the AFOSR mini-grant program. The nature of this research activity requires that this report be structured into several, partiallv autonomous subparts. The first subpart provides a general introduction. The next 3 parts describe a series of empirical analyses and investigations. The last part provides a summary and conclusions.

1. INTRODUCTION

Background Material

The research activity of the principal investigator during the SFRP focused on examining the Organizational Development efforts of the Air Force as conducted by the Leadership and Management Development Center (LMDC). A major mission of LMDC through its Directorate of Management Consultation is to deliver on-site behavioral consulting assistance to managers in the Air Force. The typical mode of effort is for LMDC to deploy a consulting team or teams to a particular Air Force unit at the invitation of the unit commander. On the first visit, the consulting teams attempt to identify and diagnose the causes of any apparent behavioral problems (e.g. low morale, meaningless or boring jobs, lack of rewards and recognition, poor communications, etc.) based on two methods. First, the consultants may make personal observations and conduct interviews with key personnel on-site. Second, a diagnostic survey questionnaire called the Organizational Assessment Package (OAP) developed by Hendrix and Halverson (1979) is administered to a sample of personnel in each work group. The consultants then return to LMDC and the data from the OAP are analyzed and interpreted along the dimensions listed in

Table 1. Scores are then compared to Air Force norms. Using these data, along with their own subjective interpretations, the consulting teams plan a return visit to the site (i.e. about 6-8 weeks following the first visit).

On the return visit, a variety of actions (i.e. interventions) are possible. The basic action is for the consultants to provide the supervisors of work groups with aggregated OAP data for their own work groups. To protect the anonymity of individual respondents, these data are provided only to groups of size four or greater. The data are presented in the form of group averages on the various OAP dimensions and are compared with Air Force norms developed for groups with similar functions (e.g. aircraft maintenance, materials and supply, civil engineering, etc.). The data may be presented in a variety of ways. At the least, the work group supervisor is given a summary statement of the data along with a jointly developed action plan for eliminating any problems. In other circumstances, the consultant may meet with a group of supervisors, one-on-one with an individual supervisor, or with a supervisor and his/her work group to provide and explain the feedback and jointly develop a management action plan. Hence the intensity with which feedback and assistance are provided may vary and should follow a trend where the work units that are lowest relative to the norms on the OAP receive the most intense treatment. It is the logical preference of the consulting teams, given limited time and manpower, to spend the most time with the groups who need it most. In addition to providing feedback to the supervisors, feedback packages are also given to officers higher in the hierarchical structure for all work groups under their command. Specifically, they are typically provided with the OAP feedback packages for all of the groups under their command.

Besides for providing feedback, the consulting teams, at LMDC's discretion, may provide additional assistance to supervisors, workgroups, groups of supervisors or groups of workgroups. This assistance ranges from formal presentations on topics such as motivation, delegation, situational leadership, communication, recognition and values to workshops and exercises involving team building, problem solving, communication, conflict resolution and group decision making. Again, these more intensive forms of consultation are generally reserved for a small number of groups who seem to particularly need help.

Three to six months following the second visit, the OAP is again administered to the groups. This is done to see if the situation has improved and can indicate needed follow-up actions. The second administration may also be used to evaluate the effectiveness of the various intervention techniques and, thus, permit LMDC to fine-tune their efforts.

In the terminology of the O.D. literature, the Air Force's methods may be described as a hybrid form of the survey feedback technolomy (Nadler, 1977). Survey feedback, as an organizational change technology, was popularized by the Institute for Social Research of the University of Michigan during the late 1950's. Since that time, studies have indicated that the technique, if applied correctly, may have a positive effect on worker attitudes (Miles, Hornstein, Callahan, Calder and Schiavo, 1969; Mann, 1969; Bowers, 1973; Brown, 1972). The findings are equivocal, however, on the behavioral and performance related results of survey feedback. For example, Miles et al. (1969) found that few of the actions or overt changes discussed in the feedback meetings were ever implemented. The Air Force approach differs from the pure survey feedback approach in its use of additional O.D. devices to implement and/or encourage specific behavioral changes.

SFRP RESEARCH — The author's SFRP project centered on using the OAP data to evaluate the impact of the consultation effort (Conlon, 1980). Several questions were asked. First, did the consultation effort produce changes on the OAP indicators? Second, if changes were found, were they contingent on supervisory characteristics, work group cohesiveness and work group size. Specific hypotheses were formulated about the latter question.

The first question required a fairly complex methodology. First, a factor analysis was used as a guideline to construct four indicators from the existing scales of OAP which would cover a range of organizational activity. These indicators were (1) supervisory characteristics, (2) organizational climate (morale), (3) perceived group productivity and (4) task motivating potential score (MPS). These scores were constructed by unit weighting the items, summing responses and dividing by the total number of items included on the scale. The existing data base contained a total of over 50000 cases, but pre and post intervention data were available on about 8000 cases. Because observations were uniquely identified for matching purposes only at the group level, data were aggregated yielding about 443 groups of greater than three people on which the change from before to after the intervention could be assessed.

The problem with evaluating change was that the experimental design in the existing data was (is) a multiple treatment (i.e. consultation type), pre/post design of the form:

$$
\begin{array}{lll}
O_B - T_1 - O_A \\
O_B - T_2 - O_A \\
\vdots \quad \vdots \quad \vdots \\
O_B - T_n \quad O_A
\end{array}
$$

where: O_B - is the OAP pre-treatment measure

O_A - is the OAP post-treatment measure

T_i - is treatment type i.

n - is the total number of treatments

This is a variant of the single group pre/post design which, according to Cook and Campbell (1979, pp. 99-103), is very common in evaluation studies but has three serious drawbacks. One drawback is the history confound; that is, the change from O_A to O_B could be due to events extraneous to the treatment. In the present multiple groups design, this threat is lessened. The second confound is the regression artifact or regression toward the mean effect. Briefly, because of the joint occurence of nonrandom assignment of treatment to groups and measurement error, groups scoring lowest on the pre-test (which receive the most intense treatment) will show the most positive change based only statistical artifacts. Likewise, groups scoring highest on the pre-test will show negative change. Our analysis of the OAP data clearly demonstrated the presence of this effect. Finally, a maturation confound is possible. That is, growth and development of respondents over time will affect change scores. This was not a plausible problem in the present context. Hence, the regression artifact was the major methodological threat to valid interpretation of the data and, given the absence of control groups in the design, had to be dealt with statistically.

In order to assess whether the consultation effort had an effect, the regression artifact had to be separated from any true treatment effects. Normally, this would be done using a non-treatment equivalent control group. In the absence of such a group, statistical means were used. It was assumed that the intensity of consultation efforts would systematically vary as a

function of the pre-score. Specifically, anecdotal evidence suggested that, because of time and resource limitations, consultants spent far less time and energy on groups that were near or above the norm on pre-test measures. This suggested that the amount of positive change resulting from treatment should decrease as the pre-score increased. By mapping this relationship on the regression artifact, a model was hypothesized that would separate true consultation effect from the regression effect. Figure 1 illustrates the model. A pure regression artifact would generate a linear relationship between the pre-score and the post-score that would be symmetric around the mean of the pre-score and this is illustrated by the dotted line. The assumption of differing intensity, however, would generate a curvilinear relationship between the scores such as that illustrated by the scatter of points on Figure 1. This would lead to a difference in the slopes of the linear relationships between the scores below and above the means as illustrated by the solid lines; that is, the slope of the line below the mean should be smaller than that above the mean. If the lines had different slopes, it would be evidence of a consultation effect.

The model was tested for the four measures by performing separate regression analyses of the post-score on the pre-score for observations below and above the pre-score mean. The results of these analyses are presented in Table 2. The slopes were found to be different in the expected direction and statistically significant (i.e. $p < .05$) for the organizational climate and group productivity scales. Although the slopes were different in the expected direction, the results for the supervisory characteristics and motivating potential scales were not significant. From this analysis it was concluded that consultation, at least, affected perceived group productivity and climate. The failure to obtain results for the motivating potential score was

not surprising given the nature of the measure. Task motivation is most affected by structural changes in the way work is done. We would expect this measure to be especially susceptible to change through job enlargemnt or enrichment. It is not clear that survey feedback should have an effect. In contrast the failure to identify change in supervisory behavior is more troublesome. It is possible that survey feedback caused a degree of defensiveness and rigidity on the part of the supervisors, hence they failed to accept the fact that they needed to change. This possibility should be studied further.

The second part of the SFRP project involved the analysis of moderator effects. Three hypotheses were tested. First, it was hypothesized that characteristics of the supervisor would affect the implementation of change. The rationale for this hypothesis was that since the consultants had only a limited time to spend on each unit, they often used the supervisor as a conduit for feedback and depended on him/her for the implementation of corrective actions. In order to study this hypothesis, the leadership items were decomposed to three scales. General Communication consisted of items that described a leader's propensity to communicate expectations and feedback. Initiating Behavior was constructed from items that described the degree to which the supervisor established work procedures, explained procedures, set specific goals and facilitated performance improvements. Finally, Bilateral Communication measured the frequency of two-way communication between supervisor and subordinates. Each of these were used to predict change scores.

Second, group cohesiveness was hypothesized to affect change. The more a group worked together cooperatively, the more likely the implementation of change. This scale was created from items measuring the extent of teamwork and satisfaction with co-workers in the work group.

Finally, group size was expected to be inversely related to implementation of change. The arguement was that the larger the group, the more formal the group procedures and the more resistance to change that would exist. Group size was measured by the number of persons in a work group at the pretest.

Each of these hypotheses was tested by regression the post-score of each of the criterion measures on its pre-score plus the moderator. In equation form, this is:

$$X_n = a + b_1 v_n + b_2 m + e$$

where: X_n = Post-score on measure N

V_n = Pre-score on measure N

m = Moderator variable

b_1 = Regression coefficient (beta)

e = Error

a = Intercept

This technique is equivalent to the analyses of covariance and is a preferred technique for analyzing data from non-equivalent groups (Kenny, 1975; Markus, 1979; Reichardt, 1979).

The regressions yielded significant results only for the supervisory hypothesis. The group cohesiveness and group size hypotheses were not supported. The results of the supervisory hypotheses are reported in Table 3. None of the supervisory measures predicted the supervisors score which is not surprising since a composite supervisory score had already been partialled from the post-score to create a residual change score (i.e. the non-equivalent groups adjustment). The bilateral communication index predicted the organization score change. All three predicted group score change. Finally, none

predicted change in MPS which may suggest that the supervisor can do little to improve an inherently poor task.

These results were very consistent with our previous analyses and suggested that consulting affected change only for the climate and perceived productivity scales. Secondly, it suggested that the unit supervisor may be a key individual for making change work. The picture that emerges from the strong bilateral communication result is that the supervisor who periodically holds meetings with the group for the purpose of discussing problems and generating solutions is the most effective user of the data provided by the consultation.

This notion was further assessed by examining the relationship between change and a behavioral report by group members of whether the supervisor discussed the feedback with the group after the intervention (i.e. variable X516). This variable was measured concurrently with the post-measures of the criterion variables and, therefore, relationships with the criterion could be inflated by "methods" variance. The criteria were all regressed on variable X516. The results are presented in the last column of Table 3. As is obvious, this variable is strongly related to change on all scales which suggests that the success of the survey feedback methodology is highly dependent on the ability of the supervisor to pass the feedback along to his/her subordinants and to use it, bilaterally, to solve problems.

The final step in this analysis was to relate the feedback measure (X516) to the pre-score supervisory items. Specifically, we hypothesized that we could predict whether a supervisor would use the feedback from the time one measures. The Pearson Product-Moment correlations between the supervisory items and X516 were calculated. The results, indicated significant relationships between X516 and all the supervisory items. A follow-up stepwise

regression indicated that variables measuring setting specific goals and holding group meetings provided the best possible prediction of X516 with a multiple correlation coefficient of 0.33 ($F = 23.05$, $P < .001$).

The emergent picture was clear. The success of the data feedback technology appeared to depend largely on the on-going behavior of the supervisor who must interpret the data for subordinates and take action. The most likely supervisor to do this is one who, already, habitually sets specific goals with subordinates and holds group meetings to exchange information, set goals and solve problems. These results are entirely reasonable in light of the resource constraints placed on DMC in terms of the limited attention that can be paid to each work group.

The Structure of This Report

The remainder of this report contains three sets of empirical analyses which are sectioned as follows. Section two reports the restructuring of the OAP into a new set of subscales (i.e. factors) using principal components analysis with varimax rotation. Section three investigates the impact of consultation using a quasi-experimental design. Section four examines variance in the consultation effort. Finally, a discussion section ends the report.

2. DEVELOPING SCALES

The original development of the OAP by Hendricks (1979) and Hendrix and Halverson (1979a; 1979b) was, of necessity, limited in the size and scope of the sample used to estimate its validity and reliability. Because the sample data available for this research were collected from a considerably larger population, a decision was made to construct a new set of scales based on a

factor analyses of the new sample data. In addition, the internal consistency and test-retest reliability of the scales was also assessed.

Principal components analysis with varimax rotation was applied to the OAP scale items using the sample of respondents on which it was possible to match pre-feedback with post-feedback OAP responses (N = 2083). The result of this analysis was reasonably consistent with the previous structure of the OAP. Using the rule that eigenvalues should exceed 1.0 for factors to be retained, 14 factors were identified which cumulatively accounted for about 68% of the variance in the data. Following a varimax rotation, only the first nine of these had items loading in excess of 0.4. The remaining factors were characterized by the absence of any large loadings or distinctive patterns, hence they were dropped from further consideration.

Table 4 presents the results of this analysis in terms of the nine new factors which emerged, the percentage of variance captured by the factor, the number of items loading above .4 on the factor, the internal reliability of the factor as measured by coefficient alpha, the test-retest reliability of each factor (i.e. composed of summed, unit weighted items) as measured by the correlation between the time 1 and time 2 measure of the factor within the no-treatment group (n = 383), and the OAP measures which comprise these factors.

The major change from the original structuring of the OAP resulting from this reanalysis is the dropping of 12 items from inclusion in any scale, and the reduction in the number of factors from the original 19 non-redundant factors to 9 factors. Obviously, the impact of this change would be to reduce the breadth of constructs which the OAP can claim to measure. Strict psychometric criteria, as they are being applied here, suggest that the number of independently measured constructs is no greater than nine and that any other

arrangement of items into scales would enable a variety of misinterpretations of the data.

In general, the internal reliabilities obtained for the scales were quite high, especially for those factors which have a small number of items. The test-retest correlations were less impressive, the largest being .64 and the majority being less than .40. The exact reasons for this instability over time are unclear and can only be speculated on. One possible reason for these was the relatively long period of time between administrations, about six months, and the variety of other "treatments" that could have occurred in those periods including changes in command, changes in co-workers and changes in budgets, goals and structures. The data did not permit us to adequately account for such factors in our analysis, hence they inflate the error term both in our scale reliability calculations and in our analyses of treatment effects. Nonetheless, in comparison with the data used in other similar longitudinal field analyses of consultation effects, the available data seemed adequate for proceeding with further analyses.

3. ASSESSING THE IMPACT OF CONSULTATION

The major objective of this research was to investigate the impact of the consultation by LMDC terms on the client unit through changes in OAP indicators occuring from time 1 to time 2. It was possible to assess these changes using the experimental design outlined in Figure 2. The control groups consisted of 380 individuals from a single functional area who had responded to the OAP as part of a special project but had not recieved any consultation. The experimental group consisted of the 92 respondents who were in the pool of cases for which pre-measures could be matched with post-measures. The data from these groups could be used to form a non-equivalent

control group, quasi-experimental design (c.f. Cook and Campbell, 1979; p. 103). Because the selection of respondents into each group was not the result of a purely random process, the design was only quasi-experimental. There are two major threats to the internal validity (i.e. accuracy of causal inferences) of such a design. First, it is possible that the two groups could be different in a way that would create differential changes from time 1 to time 2 regardless of treatment. For example, if maturation was a possible cause of changes in indicators over time and the two groups were not equal in their rates of maturation, then the maturation effect would create a differential change from time 1 to time 2 that could either mask the impact of the treatment effect (i.e. where maturation is greater in the control group) or be confused with a treatment effect (i.e. when maturation is greater in the experimental group). These biases are called selection interactions. The second possible bias is the well known regression toward the mean effect. This effect refers to the trend among non-error-free indicators to equilibrate toward their true scores over time. Hence a group that scores "high" at time 1 may have scored high partly because of random error and will probably score lower at time 2.

It is difficult to address the selection interaction bias for these data because little is known about how differences in the characteristics of groups would affect the indicators over the 6 month period between administrations of the OAP. The potential for such biases can best be examined by reviewing the selection process. The consultation process begins with the invitation by a commander for LMDC teams to come to his/her unit. One might ask why a commander would invite them. One reason may be concerns or suspicions about unit effectiveness and a felt need for outside help and support. A very different reason might be a desire for external validation of felt excellence of the

unit. We cannot know the exact motives of the commanders for our experimental and control groups, but we do know that in all cases the LMDC teams were invited based on the same premise, that is, that they had some expertise that could potentially assist unit commands. The only point that differentiated the experimental from the control groups in this regard was the objective of the consultation. The experimental data were part of the normal LMDC consultation mission. The control group data were gathered as part of a special project conducted for a particular command which, in the case of the 380 respondents, had no treatment by plan. In summary, although the groups were not formed by a true random process, the selection of respondents into each of the two groups occurred through the same mechanisms. For that reason, and because the groups were not substantially different in their prescores (see the analysis presented below), the investigators feel that the potential for selection interactions is low.

The regression problem was somewhat more managable. As a first step in evaluating the potential for regression biases, the two groups were compared on prescores across the 9 factors. This comparison is presented in Table 5. Only two for the factors, organizational climate and adequacy of job resources were significantly different between the groups. The apparent similarity of the two groups on most factors substantially reduced our concerns about large differences in selection criteria and reduced, to some extent, the regression threat.

In spite of the similarity of the groups, it was decided that analysis of covariance (ANCOVA) would be an appropriate measure to further minimize regression problems. It should be noted, however, that ANCOVA is at best "First aid" and cannot be relied on to eliminate all of the problems caused by non-randomness (Cook and Campbell, 1979).

The Treatment Variance Problem

In most experiments, great care is taken to minimize sources of variance within the various treatment and control conditions because it will inflate the "within cell" or error variance in the analysis and raise the possibility of failing to reject the null hypothesis when it should be rejected. In this study, because the treatment was (1) administered by a variety of consultants and (2) depended on the clients for a good bit of follow-through, there was undoubtedly uncontrolled variance in the treatment. In the SFRP, it was assumed that the variance was conditional on pre-score levels of the factors. In the present analysis, steps were taken to try to account for or control such variance.

Variable X516, the respondent's reports of how much feedback and discussion there was of the survey data in their work groups, was used as an indicator of treatment intensity. The measure was trichotomized into three levels where 1 or 2 was low feedback, 3, 4 or 5 was medium and 6 or 7 was high feedback. This led to the creation of 4 experimental conditions which were ordered on a single experimental factor, intensity of survey feedback treatment. The control group, which got absolutely no feedback, was the no treatment condition. The three levels of feedback, low medium and high, were the remaining 3 conditions.

Analysis and Results

The analyses proceeded in two steps. The first step involved multivariate analyses. The correlation matrix presented in Table 6 indicated substantial correlations between some of the factors (i.e. dependent variables), thus requiring the utilization of multivariate techniques. The second step involved the interpretation of multivariate effects using correlations and univariate ANCOVA's.

lmlco/10/6/82

Multivariate Analyses. A one-way multivariate analyses of covariance was used on standardized values of the nine factors across the four experimental conditions indicated above. The result was a significant multivariate effect (Wilks λ = .908, s = 3, m = 2.5, N = 224.5, df = 27, $p < .03$) across the four cells. The correlation of the dependent measures with the significant canonical variate, which are useful in the interpretation of the measure (Borgen and Seling, 1979) are presented in Table 7. These correlations are interpretable as indicating the extent to which each dependent measure is being impacted by the consultation. The pattern of correlations suggests that the consultation had by far the largest impact on perceptions of supervisory behavior (r = .667) and little or no impact on perceived climate (r = .074). It had a moderate impact on task perceptions, perceived autonomy, perceived goal quality and perceived resource adequacy and a small impact on attitude toward enriched tasks and perceived group effectiveness.

Univariate Analyses. Following the demonstration of a multivariate effect for consultation, it is possible to further investigate the effect through univariate analyses. Table 8 presents the results of the univariate ANCOVA's. The covariates for these analyses include all of the prescores for all 9 dependent measures. Four of the univariate effects were significant, those for perceptions of supervisory behavior, perceptions of tasks, perceived opportunity for career development and goal quality. These effects could be further examined looking at specific cell means. The conditional means for each of the variables with significant univariate ANCOVAs are presented in Table 9. The general trend for all measures is for the high and medium feedback means to dominate the means in the low and no feedback conditions. In addition, the mean in the low feedback condition is slightly less than that in

the no feedback condition on the perceived task characteristics and the opportunity for advancement measures. It is particularly clear that respondents must perceive at least a moderate amount of feedback and discussion in order for significant change to take place.

4. Prescore Levels, Treatment and Effects

In the SFRP, it was assumed that the intensity of treatment would vary as a function of pre-score values and that, in turn, would impact the amount of change taking place from time 1 to time 2. In the present analysis, it was possible to use the control group to estimate the "true" regression effect and, therefore, to examine the validity of the SFRP model (see section 1).

In order to investigate the regression model used in the SFRP, the data were again cut at the (sample) mean of the prescore for each factor. Four sets of regressions of the change score (i.e. the dependent measure) on the pre-score were performed on each factor, two using control group data above or below the mean, and two using experimental group data above or below the mean. The regression coefficients, presented in Table 10, were examined for trends.

It is obvious that the trends did not conform to expectations. If the treatment was more intense the lower the pre-score, then the slope of the experimental group regression below the mean should be greater than that of the control group. On average, and in 8 of the 9 measures, this was not the case (i.e. all except "goal quality"). In fact, it seems that the opposite was occuring, that is, slopes were steeper in the control group. The trends for data above the mean were also contradictory to the SFRP assumptions. We expected the slopes to be substantially equal, but the trend indicated that

the slope in the control group was again greater than that in the experimental group.

Based on this result, it was decided that the relationship of variable X516 (i.e. perceived feedback) to the pre-score should be examined. As a first step, the correlations between the pre-scores and X516 were investigated. These are presented in Table 11. The results indicated that significant, although weak, linear relationships existed between variable X51b and five of the nine prescore measures. It is noteworthy that the five prescore measures were also moderately interrelated and appeared to form a cluster separate from the remaining four measures. Because of the possibility of a non-linear relationship, it was decided to examine the trends in X516 as a function of prescores more precisely. This was done by partitioning the pre-score values of the four measures on which significant change was found (see Section 2) into octants. These cuts were made to create eight groups with relatively equal cell sizes, although this goal was partially achieved at best. The means of variable X516 were then calculated and analyzed for significant trends across the eight cells. The results of these analyses are presented in Table 12. Eta^2 in this table refers to the approximate percentage of the total variation in variable X516 that could be explained by the prescore measures.

The analysis revealed that the relationship between prescores and variable X516 was not linear. It was apparent that the perceived amount of feedback given was greatest for the largest values of the prescore measures. The trend, however was not generally linear across all values of the prescores. In prescore variable 903, for example, the mean values of variable X516 were greatest in the first and eighth octants (i.e. the extremes). In all of the

prescores, the mean of variable X516 appeared to "dip" somewhat in either the fourth or fifth octant and to increase steadily after that point.

These trends suggest that the process or processes that link prescore values to feedback may be more complex than expected or hypothesized in the SFRP. In particular, the data may indicate the mediation of at least two processes, one which accounts for feedback being given to the "worst" groups, and another accounting for feedback in the best groups. Although sufficient data were not available for further elucidation, it is possible that for the groups with the smallest prescores, the consultants engaged in the most intensive consultation thus accounting for the "peaks" observed, depending on the particular prescore variable, in one of the first three octants. The other "peak," generally occuring in octants 7 and 8, could have been caused by a very good supervisor who was able to use the survey data as reinforcing feedback for his subordinates thus enhancing their scores on the OAP. The mid-range octants (i.e. 4 through 6) could suffer from a combination of (1) less attention from the consultants and (2) incomplete follow-through by the supervisor. It should be understood that these "explanations" are purely speculative and, at best, may serve to motivate and direct further investigation.

5. DISCUSSION

The investigations reported here suggest a variety of conclusions.

(1) The OAP, when subjected to principal components factor analysis, yields fewer factors than are assumed in the present version.

(2) The internal consistency of the factor-analytically derived scales is good as indicated by the coefficient alpha statistic.

(3) The intertemporal reliabilities of most of the scales are poor. This is probably inevitable because of the relatively long time span between measurements and the multiplicity of factors,

besides consulting interventions, which could change the scale values over time.

(4) There is evidence that the intervention effort has had some effect on at least four of the nine factor analytically derived scales. These scales are not independent (i.e. uncorrelated), hence the effect should not be thought of as four separate and distinct phenomena.

(5) The relationship between prescores, consulting effort and change appears to be complex and really cannot be satisfactorally studied with the data-on-hand. We feel, however, that the key to refining and imporving LMDC's efforts is a better understanding of what types of interventions work best in a given situation, and why. Such issues can only be addressed through controlled experimental designs conducted within the Air Force context.

Suggested Directions

At this point, given the results of the SFRP and the research reported here, several issues should be considered. The first concerns what can or should be done with the existing evaluation data. The second concerns the status of the evaluation effort as it is presently conducted. Finally, should the focus and/or design of the effort be substantially changed in the future?

The existing data - The present research attempted to utilize the existing data base to conclude as much as possible about the effectiveness of the consultation efforts not only in terms of overall effects but trying to ascertain the processes leading to effects. These efforts were only partially successful. The multivariate analysis of covariance on the OAP indicators indicated some changes in the OAP indicators that were attributable to the experimental design. The unreliability of the data over time, however, makes the detection of effects difficult and prone to conservative error (i.e. failure to detect real effects).

Our follow-up analysis of the consulting process, that is, attempting to identify and predict treatment variance, resulted in the detection of a curvelinear relationship between prescore measures and perceived feedback that

is subject to lots of speculative explanations, none of which are based on strong empirical verification.

This investigator concludes that the further use of the experimental designs presently usable with the existing data base to ascertain the causal properties of the LMDC efforts would be futile and unwise. The restricted availability of control groups (i.e. encompassing one functional area), which greatly limits sample size and the presence of uncontrollable treatment variance, constrain the investigator's options for making strong inferences about the data.

The status of evaluation - Given the above assertions about the data, the investigator recommends a modification of the evaluation effort as is not likely to yield more than is already known about LMDC consulting efforts. Further, the use of the current procedure for evaluating the ongoing impact of consultation on a workgroup basis for feedback purposes is questionable given the regression effects and the dangers of trying to estimate and control for such effects based on limited control group data. I encourage LMDC to seriously weigh the benefits of continuing the present evaluation effort against its cost and to consider alternative evaluation strategies.

Alternative evaluation methods - As an alternative to the present evaluation methodology, it is suggested that LMDC consider a series of controlled studies which would investigate, by plan, the effectiveness of particular types of consultation efforts across various functional specialties in the Air Force. For example, one could use the existing data base to identify for particular functional specialities (e.g. chaplains, security police, civil engineers, flight line mechanics, etc.) those areas, according to the OAP, where they are "deficient". By plan, then, a variety of consulting approaches could be attempted and evaluated on those groups as part of the ongoing

consultation process in order to ascertain the effectiveness of each aproach. Each such "experiment" would require the cooperation of the consultants and the client group, especially with regard to gathering control group data. Designs are available, however, that would permit the collection of valid data and the delivery of quality consultation simultaneously.

I suspect that the above suggestion will be viewed as costly and troublesome when compared with the present evaluation methodology. However, if one accepts that we have learned all we can from the present evaluation effort and that it is of limited usefulness as a feedback device to the consultants, then the choices are to either restructure the evaluation so that we can learn from it and thus "fine tune" the consultation effort, or terminate the evaluation aspect of LMDC operations. This investigator feels that the evaluation funds and energies would be best spent on studies which are limited in scope but high in information yield as compared to the present method which is broad in scope but is not likely to yield very much additional information of use to the Air Force.

lmlco/10/6/82

REFERENCES

1. Borgen, F. and M. J. Seling. "Uses of Discriminant Analysis Following Manova: Multivariate Statistics for Multivariate Purposes." Journal of Applied Psychology, 1978, 63, 689-697.

2. Brown, L. D. "Research Action: Organizational Feedback, Understanding and Change." Journal of Applied Behavioral Science, 8, 1972, 697-704.

3. Conlon, E. J. "Investigations of Behavioral Consultation in the Air Force." Technical Report prepared for the 1980 USAF-SCEEE Summer Faculty Research Program, AFOSR, Bolling AFB, D.C. 20332, 1980.

4. Cook, T. D. and D. T. Campbell, Quasi-Experimentation, Chicago, IL: Rand-McNally, 1979.

5. Hendrix, W. H. "Organizational Assessment Indices of Effectiveness" (AFHRL-TR-79-46). Brooks AFB, TX: Air Force Human Resources Laboratory, 1979.

6. Hendrix, W. H. and V. Halverson. "Organizational Survey Assessment Package for Air Force Organizations" (AFHRL-TR-78-93). Brooks AFB, TX: Air Force Human Resources Laboratory, 1979(a).

7. Hendrix, W. H. and V. Halverson. "Situational Factor Identification In Air Force Organizations (AFHRL-TR-79-10). Brooks AFB, TX: Air Force Human Resources Laboratory, 1979(b).

8. Kenny, D. "A Quasi-Experimental Appraoch to Assessing Treatment Effects in the Non-Equivalent Control Group Design," Psychological Bulletin, 1975, 82, 345-362.

9. Mann, F. C. Studying and Creating Change. In W. Bennis, K. Benne and R. Chin (eds.), The Planning of Change (2nd ed.), New York, NY: Holt, Rinehart and Winston, 1969.

10. Markus, G. B. Analyzing Panel Data, London, U.K.: Sage Publications, 1979.

11. Miles, M. G., H. A. Hornstein, D. M. Callahan, D. M. Calder and R. Schiavo. "The Consequences of Survey Feedback: Theory and Evaluation." In W. G. Bennis, K. D. Benne and R. Chin (Eds.) The Planning of Change (2nd ed.). New York, N.Y.: Holt, Rinehart and Winston, 1969, 456-68.

12. Nadler, D. Feedback and Organization Development: Using Data-Based Methods. Reading, MA: Addison-Wesley, 1977.

13. Reichardt, C. S. "The Statistical Analysis of Data from Nonequivalent Group Designs," in T. D. Cook and D. T. Campbell, Quasi-Experimentation, Chicago, IL: Rand McNally, 1979, 147-206.

lmlco/10/6/82

Skill Variety
Task Identity
Job Feedback
Work Support
Need For Enrichment
Job Performance Goals
Pride
Task Characteristics
Task Autonomy
Work Repetition
Desired Repetitive Easy Tasks
Job Influences
Advancement/Recognition
Supervisory Style
Supervisory Communication Climate
Organizational Communications Climate
Work Group Effectiveness Work Interferences
Job Related Satisfaction
Job Related Training
Organizational Climate
Task Motivational Potential

Table 1

Factors Assessed By the Present Version of the OAP

Scale	Beta (Slope) For Below The Pre-Test Mean	Beta (Slope) For Above The Pre-Test Mean	Z Score Of Difference	P Level
Supervisory Characteristics	.466	.574	-	N.S.
Organizational Climate	.409	.847	3.55	.01
Group Productivity	.259	.586	2.10	.05
Motivating Potential	.596	.655	-	N.S.

Table 2

Summary of Comparisons of Slopes on Each Scale
For Below And Above The Pre-Test Means

Supervisory Measures
at Time 1

Partial r F-Value	General Comm	Initiating Behavior	Bilateral Comm	Behavioral Feedback (X515)
Supervisory Style Score	.015 0.80	-.060 1.29	.090 3.09	.323 43.73***
Organizational Climiate Score	.049 0.91	.013 0.06	.120 5.38*	.342 49.82***
Group Productivity Score	.102 3.94*	.136 6.98**	.164 10.23**	.249 24.83***
Task MPS Score	.046 0.78	.033 0.41	.070 1.84	.383 64.64***

*Indicates P < .05

**Indicates P < .01

***Indicates P < .005

Table 3

Tests of the Supervisory Moderators and the Behavioral Feedback Measure

#	LABEL	% of Variance Explained	No. of Items	RELIABILITY Internal	Test-Retest
901	Supervisory Characteristics (V404 Thru V445, V206, V278)	(35%)	19	.97	.31
902	Climate (V300 Thru V318, V241, V711, V719, V723)	(7%)	23	.95	.54
903	Task Characteristics & Pride (V201 Thru V203, V210 Thru V212, V215, V275, V705)	(5%)	9	.88	.61
904	Attitude Toward Enrichment (V249 Thru V258)	(3%)	7	.63	.25
905	Perceived Group Effectiveness (V238, V259, V260, V261, V264, V265, V312, V709)	(3%)	8	.88	.39
906	Task Autonomy (V213, V214, V270, V271)	(2.5%)	4	.80	.64
907	Oppt'y. For Career Development (V234, V239, V240, V241, V276)	(2%)	5	.79	.50
908	Goal Quality (V217, V221, V273, V274)	(2%)	4	.78	.35
909	Adequacy of Job Resources (V 07, V277)	(1.5%)	2	.81	.39

Table 4

FACTORS DERIVED FROM PRINCIPAL COMPONENTS ANALYSIS

(VARIMAX ROTATED)

lmlcotb/10/7/82

#	FACTOR	MEANS Control (N = 380)		Experimental (N = 92)
901	Supervisory Description	86.58		91.43
902	Climate	84.43	*	90.79
903	Task/Pride	35.53		37.75
904	Attitude About Enrichment	17.90		17.35
905	Preceived Group Effectiveness	36.49		37.20
906	Task Autonomy	10.56		11.53
907	Opportunity for Career Devel.	18.69		18.95
908	Goal Quality	17.97		18.75
909	Adequacy of Job Resources	7.55	*	8.80

Table 5

A COMPARISON OF EXPERIMENTAL
AND CONTROL GROUPS ON
THE PRESCORES

(* indicates a statistically significant
difference ($p < .05$))

lmlcotb/10/7/82

	Factor #									
	901	902	903	904	905	906	907	908	909	
901	1									SUPERVISOR
902	.508	1								CLIMATE
903	.397	.626	1							TASK
904	.097	.103	.188	1						ATT. TOWARD ENRICH.
905	.479	.618	.545	.144	1					PCUD. EFFECTIVENESS
906	.335	.532	.638	.157	.432	1				AUTONOMY
907	.490	.687	.592	.111	.516	.506	1			OPPT'Y CAREER DEV.
908	.385	.515	.535	.173	.399	.382	.502	1		GOAL QUALITY
909	.263	.465	.287	.023	.261	.351	.345	.358	1	RESOURCE ADEQ.

Table 6

CORRELATIONS AMONG DEPENDENT VARIABLES

(n = 472)

lmlcotb/10/7/82

	FACTOR	CORRELATION WITH CANONICAL VARIATE
901	SUPERVISOR	.667
902	CLIMATE	.074
903	TASK	.492
904	ATT.TOWARD ENRICHMENT	.302
905	PCUD. EFFECTIVENESS	.221
906	AUTONOMY	.473
907	OPPT'Y CAREER DEVEOPMENT	.259
908	GOAL QUALITY	.498
909	RESOURCE ADEQUACY	.378

OVERALL: WILKS λ = .908 ($p < .03$)

Table 7

EFFECTS IN THE MANCOVA

(STRUCTURE MATRIX)

Dependent Measure	Hypothesis Sums of Sq.	Error Sums of Sq.	MSH	MSE	F
901	10.08	424.51	3.36	.92	3.6*
902	3.94	362.39	1.31	.79	1.7
903	7.14	326.92	2.38	.71	3.3*
904	4.21	441.23	1.40	.96	1.5
905	3.40	391.88	1.13	.85	1.3
906	3.70	324.66	1.23	.71	1.7
907	10.37	366.52	3.45	.80	4.3*
908	8.25	417.08	2.75	.91	3.0*
909	5.68	398.37	1.89	.87	2.2

Table 8

Table of Univariate ANCOVAs
(Following Multivariate Analysis)
(df = 3, 459)
(* indicates $p < .05$)

Dependent Measure		High Feedback Condition	Medium Feedback Condition	Low Feedback Condition	No Feedback Condition
901	Perceived Supervisory Behavior	0.533	0.671	0.178	-0.071
903	Perceived Task Characteristics	0.935	0.752	-0.051	-0.047
907	Opportunity for Career Advancement	0.922	0.696	-0.243	-0.013
908	Goal Quality	1.095	0.249	0.035	-0.043

Table 9

Table of Standardized Condition Means for Those Measures Having Significant Univariate ANOVAs

lmlcotb/10/7/82

	Less Than or Equal to The Mean			Greater Than The Mean		
Group	Var.	Beta	Mean	Var.	Beta	Mean
Control	901	-.80		901	-.45	
	902	-.61		902	-.35	
	903	-.68		903	-.18	
	904	-1.00		904	-.57	
	905	-.93	-.77	905	-.48	-.51
	906	-.87		906	-.31	
	907	-.61		907	-.59	
	908	-.82		908	-.78	
	909	-.59		909	-.88	
Experimental	901	-.58		901	-.65	
	902	-.08		902	-.14	
	903	-.37		903	-.20	
	904	-.71		904	-.48	
	905	-.51	-.46	905	-.28	-.33
	906	-.25		906	-.21	
	907	-.18		907	-.32	
	908	-1.00		908	-.36	
	909	-.48		909	-.35	

Table 10

Regression Coefficients for Regressions of Change Scores on Pre Scores Cut According to Experimental Group and Position With Regard to the Pre Score Mean

Prescore Measure	Correlation With Variable 516
901	.16
902	.16
903	.29*
904	.15
905	.22*
906	.35*
907	.34*
908	.30*
909	.06

* indicates $p < .05$

Table 11

Correlations Between Prescores
And Variable X516
(Perceived Intensity of Feedback)

Prescore Variable	Prescore Octant	516 Sum	516 Mean	n	ANOVA F Value	ETA^2
	1	19	1.7	7		
	2	19	2.1	9		
	3	27	2.2	12		
901	4	21	1.5	14	1.29	.10
	5	27	1.6	16		
	6	33	2.7	12		
	7	20	2.5	8		
	8	38	3.2	12		
	1	20	2.5	8		
	2	14	1.2	12		
	3	15	1.9	8		
903	4	27	1.8	15	2.08*	.15
	5	19	2.1	9		
	6	14	1.8	8		
	7	36	2.4	15		
	8	52	3.5	15		
	1	22	1.5	15		
	2	11	1.8	6	2.06*	.15
	3	23	2.1	11		
907	4	33	1.8	18		
	5	10	1.2	8		
	6	33	3.0	11		
	7	29	3.2	9		
	8	36	3.0	12		
	1	9	1.3	7		
	2	17	1.5	11		
	3	21	1.9	11		
908	4	22	1.5	15	2.37*	.17
	5	23	2.6	9		
	6	28	3.5	8		
	7	27	2.0	13		
	8	50	3.1	16		

(* indicates $p < .05$)

Table 12

Profiles of Variable X516 (Perceived Intensity of Feedback) Across Octant Values of the Prescores for Which Signigicant Change Was Indicated

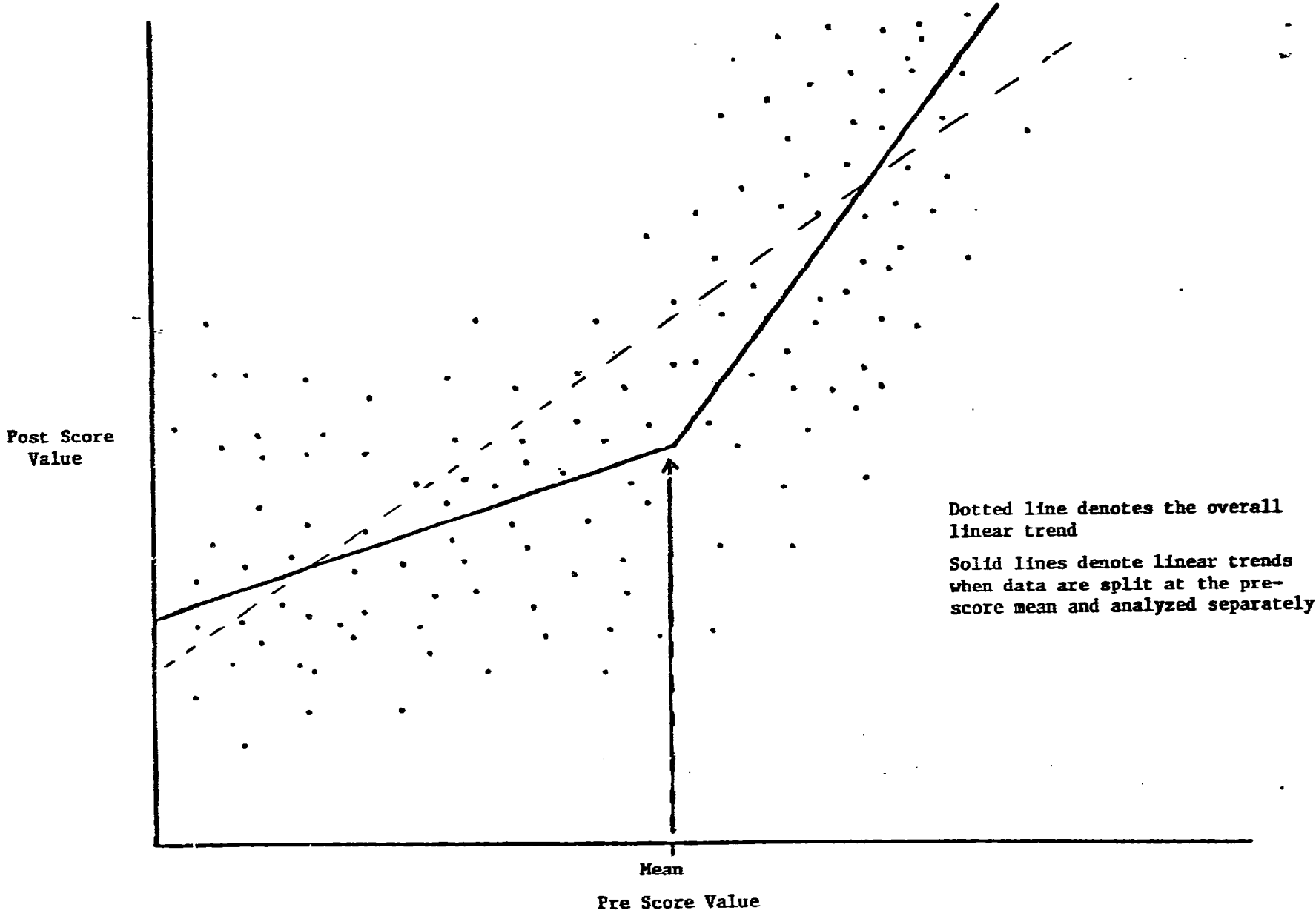


Figure 1

Hypothesized Relationship Between Pre and Post Scores

	Time 1		Time 2
Experimental Group (n = 92)	0	X (Consultation)	0
Control Group (n = 380)	0		0

Figure 2

The Non-Equivalent Control Group Design Used to Evaluate Consultation (selection was not random)

lmlcotb/10/7/82